

# Holonomic Gradient Descent and its Application to Fisher-Bingham Integral

Tomonari Sei, Nobuki Takayama, Akimichi Takemura ;  
Hiromasa Nakayama, Kenta Nishiyama, Masayuki Noro, Katsuyoshi Ohara

September 5, 2010

We give a new algorithm to find local maximum and minimum of a holonomic function and apply it for the Fisher-Bingham integral on the sphere  $S^n$ , which is used in the directional statistics. The method utilizes the theory and algorithms of holonomic systems.

## 1 Introduction

The gradient descent is a general method to find a local minimum of a smooth function  $f(z_1, \dots, z_d)$ . The method utilizes the observation that  $f(p)$  decreases if one goes from a point  $z = p$  to a “nice” direction, which is usually  $-(\nabla f)(p)$ . As textbooks on optimizations present (see, e.g., [5], [16]), we have a lot of achievements on this method and its variations.

We suggest a new variation of the gradient descent, which works for real valued *holonomic functions*  $f(z_1, \dots, z_d)$  and is a  $d$ -variable generalization of Euler’s method for solving ordinary differential equations numerically and finding a local minimum of the function. We show an application of our method to directional statistics. In fact, it is our motivating problem to develop the new method.

A function  $f$  is called a holonomic function, roughly speaking, if  $f$  satisfies a system of linear differential equations

$$\ell_1 \bullet f = \dots = \ell_r \bullet f = 0, \quad \ell_i \in D \quad (1)$$

whose solutions form a finite dimensional vector space. Here,  $D$  is the ring of differential operators with polynomial coefficients  $\mathbf{C}\langle z_1, \dots, z_d, \partial_1, \dots, \partial_d \rangle$ , and the action  $\bullet$  is defined by  $z^\alpha \partial^\beta \bullet f = z_1^{\alpha_1} \dots z_d^{\alpha_d} \frac{\partial^{|\beta|} f}{\partial z_1^{\beta_1} \dots \partial z_d^{\beta_d}}$ .

Let us give a rigorous definition of holonomic function. A multi-valued analytic function  $f$  defined on  $\mathbf{C}^d \setminus V$  with an algebraic set  $V$  is called a *holonomic function* if there exists a set of linear differential operators  $\ell_i \in D$  annihilating  $f$  as (1) such that the left ideal generated by  $\{\ell_1, \dots, \ell_r\}$  in  $D$  is a *holonomic ideal* (see [15]). The function  $f$  is called real valued when a branch of  $f$  takes real values on a connected component of  $(\mathbf{C}^d \setminus V) \cap \mathbf{R}^d$ .

We give an equivalent definition of holonomic function without the notion of the holonomic ideal ([18], [12], [15]). A multi-valued analytic function  $f$  is called a holonomic function if  $f$  satisfies linear ordinary differential equations with polynomial coefficients for all variables  $z_1, \dots, z_d$ . In other words, the function  $f$  satisfies a set of ordinary differential equations

$$\sum_{k=0}^{r_i} a_k^i(z_1, \dots, z_d) \partial_i^k \bullet f = 0, \quad a_k^i \in \mathbf{C}[z_1, \dots, z_d], \quad i = 1, \dots, d.$$

When  $n = 1$ , a holonomic function is nothing but a solution of linear ordinary differential equation with polynomial coefficients. In this case, a local minimum can be obtained numerically by a difference scheme, which is called Euler's method. Readers may think that it will be straight forward to generalize Euler's method to  $d$ -variables, which we will call *holonomic gradient descent*. However, as we will see in this paper, a generalization of Euler's method to  $d$ -variables requires to utilize the theory, algorithms, and efficient implementations of Gröbner basis for holonomic systems, which have been studied recently (see [15] and its references).

In Section 2, we will illustrate holonomic gradient descent precisely. In Sections 3 and 4, we study the Fisher-Bingham integral as a holonomic function. In Section 5, we consider problems in the directional statistics as applications of results of Sections 2, 3, and 4.

Our method is based on holonomic systems of differential equations. D. Zeilberger proposed the holonomic function approach for special function identities about 20 years ago and it has been studied in the past 20 years (see, e.g., [1] and its references). We present, in this paper, that the holonomic approach will be promising as a new method in statistics and in optimization. We note that this point of view of holonomic systems and holonomic functions has been emphasized by few literatures in statistics and in optimization.

## 2 Gradient Descent for Holonomic Functions

There are several methods of finding a local minimum of a given function  $g$ . Among them, iteration methods are the most general and are often used methods. Iterations are written as

$$z^{(k+1)} = z^{(k)} + h_k d^{(k)} \quad k = 0, 1, 2, \dots \quad (2)$$

where  $\{z^{(k)} \in \mathbf{R}^d\}$  is a sequence such that  $g(z^{(k)})$  converges to a local minimum of the function  $g$ ,  $h_k \in \mathbf{R}_{>0}$  is a step length, and  $d^{(k)}$  is called the search direction. The search direction has the form

$$-H_k^{-1}(\nabla g)(z^{(k)}) \quad (3)$$

where  $H_k^{-1}$  is a  $d \times d$  matrix. Typical choices of  $H_k$  are the identity matrix for the gradient descent and the Hessian matrix of  $g$  for Newton's method [5].

The iteration method is a numerical method. When the function  $g$  is a holonomic function, we can apply the Gröbner basis method, which is an algebraic and symbolic method, for the evaluation of the search direction. When we are given a Gröbner basis  $B$ , a set of monomials  $S$  is called the set of the *standard monomials* of  $B$  if it is the set of the monomials which are irreducible (non-divisible) by  $B$  (see, e.g., [4], [17]). Let  $g(z_1, \dots, z_d)$  be a holonomic function and we suppose that it is annihilated by a holonomic ideal  $I$ . Let  $S$  be the set of the standard monomials of a Gröbner basis of  $RI$  in  $R = \mathbf{C}(z_1, \dots, z_d)\langle \partial_1, \dots, \partial_d \rangle$ , which is the ring of differential operators with rational function coefficients. The cardinality of  $S$  is finite and is called the *holonomic rank* of  $I$ . We may suppose that  $S$  contains 1 as the first element of  $S$ . Since the function  $g$  is holonomic, the column vector of functions  $G = (s_i \bullet g \mid s_i \in S)^T$  satisfies the following set of linear partial differential equations (see, e.g., [15, p.39])

$$\frac{\partial G}{\partial z_i} = P_i G, \quad i = 1, \dots, d \quad (4)$$

where  $P_i$  is a square matrix with entries in  $\mathbf{C}(z_1, \dots, z_d)$ . In fact, when the normal form of  $\partial_i s_m$  by  $G$  in  $R$  is  $\sum_n c_{mn}^i s_n$ , the rational function  $c_{mn}^i$  is the  $(m, n)$ -th entry of the matrix  $P_i$  (see, e.g., the reduction algorithm in [17]). Note that each equation can be regarded as an ordinary differential equation with respect to  $z_i$  with parameters  $z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_d$ . We call the system of differential equations (4) *the Pfaffian system (or equations) for  $g$* . The first entry of  $G$ , which is denoted by  $G_1$ , is  $g$ .

A remarkable fact on holonomic function in this iteration scheme is that the gradient of  $g$  and the Hessian of  $g$  can be written in terms of the vector function  $G$ , which implies that we can evaluate the search direction for the gradient descent from the value of  $G$ . This fact is an easy consequence of the Gröbner basis theory, but it is fundamental for the optimization of holonomic functions. Precisely speaking, we have the following formula.

**Lemma 1** 1. Let  $\sum_{s_j \in S} a_{ij} s_j$  be the normal form of  $\partial_i = \partial/\partial z_i$  by the Gröbner basis  $B$  of  $RI$  in  $R$ . Here we have  $a_{ij} \in \mathbf{C}(z_1, \dots, z_d)$ . Let  $A$  be the matrix with entries  $a_{ij}$ . Then, we have

$$(\nabla g)(z^{(k)}) = A(z^{(k)})G(z^{(k)})$$

and

$$(\nabla g)(z^{(k)}) = ((P_1 G)_1, \dots, (P_d G)_1)(z^{(k)})$$

where  $(v)_1$  notes the first entry of a vector  $v$ .

2. Let  $\sum_k u_{ijk} s_k$  be the normal form of  $\partial_i \partial_j$  with respect the Gröbner basis  $B$  where  $u_{ijk} \in \mathbf{C}(z_1, \dots, z_d)$ . Then, we have

$$\frac{\partial^2 g}{\partial z_i \partial z_j}(z^{(k)}) = (u_{ij1}(z^{(k)}), \dots, u_{ijd}(z^{(k)}))G(z^{(k)})$$

and

$$\frac{\partial^2 g}{\partial z_i \partial z_j}(z^{(k)}) = \left( \left( \frac{\partial P_i}{\partial z_j} + P_i P_j \right) G \right)_1$$

*Proof.* Since,  $\partial_i - \sum_j a_{ij} s_j \in RI$  and  $(RI) \bullet g = 0$ , we have  $\partial_i \bullet g = \sum_{s_j \in S} a_{ij}(s_j \bullet g)$ . Then, we have the first identity of (1). Since  $\frac{\partial G}{\partial z_i} = P_i G$  and  $G_1 = g$ , we have the second identity of (1). The first identity of (2) can be shown analogously. Differentiating  $\frac{\partial G}{\partial z_i} = P_i G$  by  $z_j$ , we have  $\frac{\partial^2 G}{\partial z_i \partial z_j} = \frac{\partial P_i}{\partial z_j} G + P_i \frac{\partial G}{\partial z_j} = \left( \frac{\partial P_i}{\partial z_j} G + P_i P_j \right) G$ . Thus, the second identity of (2) is obtained. Q.E.D.

It follows from this lemma that we obtain the following gradient descent for holonomic functions to find a local minimum. We shortly call the method *holonomic gradient descent*. Note that this is a symbolic-numeric algorithm.

**Algorithm 1** (Holonomic gradient descent)

1. Obtain a Gröbner basis of  $RI$  in  $R$  and the set of the standard monomials  $S$  of the basis.
2. Compute the matrices  $P_i$  in (4) by the normal form algorithm and the Gröbner basis and the set of the standard monomials.
3. Compute the normal form  $\partial_i$  by a Gröbner basis of  $RI$  and determine the matrix  $(a_{ij})$ .
4. Take a point  $z^{(0)}$  as a starting point and evaluate numerically the initial value of  $G$  at  $z = z^{(0)}$ . Denote the value by  $\bar{G}$  and put  $k = 0$ .
5. Evaluate numerically  $(a_{ij}(z^{(k)}))\bar{G}$ , which is an approximate value of the gradient  $\tilde{g} = \nabla g$  at  $z^{(k)}$ . If a termination condition of the iteration is satisfied, then stop.
6. Put  $z^{(k+1)} = z^{(k)} + h_k \tilde{g}$ , (move to  $z^{(k)} + h_k \tilde{g}$ ).
7. Obtain the approximate value of  $G$  at  $z = z^{(k+1)}$  by solving numerically the Pfaffian system (4) by the Runge-Kutta method (see, e.g., [11]). Set this value to  $\bar{G}$ . Increase the value of  $k$  by 1. Goto 5.

Here,  $h_k$  is the step length, which should be chosen by standard recipes of gradient descent.

Let us give two notes on numerical evaluations of  $G$ . (1) The computation of the initial value  $G$  requires a method depending on a given problem. In case of the Fisher-Bingham integral, we use a numerical integration method. (2) We use the Runge-Kutta method to evaluate  $G$  at  $z^{(k+1)}$  from the value of  $G$  at  $z^{(k)}$ . Precisely speaking, we have

$$\frac{dG(c(t))}{dt} = \sum_{i=1}^d \frac{dc_i}{dt} \frac{\partial G}{\partial z_i} = \sum_{i=1}^d \left( \frac{dc_i}{dt} P_i \right) G$$

for any smooth vector valued function  $c(t)$ . We use this expression to numerically solve the Pfaffian system to the direction  $\tilde{g}$ .

Elements of  $P_i$  are rational functions. The union of the zero sets of the denominators of elements of  $P_i$ 's is called the *singular locus* of the Pfaffian equations (4). It is known that holonomic functions are holomorphic in the complement of the singular locus of corresponding Pfaffian equations. We can apply known convergence criteria to this algorithm (see, e.g., [16]) when we look for a local minimum in a connected domain in the complement of the singular locus. Hence, we have to limit the search domain of a local minimum in the connected domain.

The holonomic gradient descent can be applied to a large class of optimization problems. It is well known that when  $f$  and  $g$  are holonomic functions, then the sum  $f + g$  and the product  $fg$  are also holonomic functions. A remarkable fact is that when  $f$  is a holonomic function in  $z_1, \dots, z_d$ , then the definite integral  $\int_{a_d}^{b_d} f(z_1, \dots, z_d) dz_d$  is also a holonomic function in  $z_1, \dots, z_{d-1}$ . We have algorithms to find systems of differential equations for the sum, the product, and the definite integral. As to these topics, see, e.g., [1], [9], [10], [11], [15] and their references. It follows from these results that we can present our algorithm in the following form.

**Algorithm 2** (Holonomic gradient descent for integrals)

Input: a definite integral  $F(z) = \int_C f(z, t) dt$  with parameters  $z = (z_1, \dots, z_d)$  where  $f(z, t)$  is a holonomic function of which annihilating ideal is  $J$ .

A holonomic function  $g(z)$  of which annihilating ideal is  $J'$ .

Output: An approximate local minimum of  $g(z)F(z)$  for  $z \in E$ .

1. Apply integration algorithms for the holonomic ideal  $J$  (see, e.g., [1], [9], [10], [11], [15] and their references) to find a holonomic ideal  $\int J$  annihilating the function  $F(z)$ .
2. Obtain a holonomic ideal  $I$  which annihilates  $g(z)F(z)$  from  $\int J$  and  $J'$  (see, e.g., [18], [11]).
3. Apply Algorithm 1 for  $I$  where starting values of  $F(z)$  and its derivatives are computed by a numerical integration method.

We note that integration algorithms require some conditions for the domain of the integration  $C$ . The domain  $C$  must satisfy the conditions. For example, when  $C$  is a product of segments and  $C$  is contained in the complement of the singularities of  $f(z, t)$ , the domain satisfies the conditions. The search domain  $E$  must be in the complement of the singular locus of the Pfaffian equations for  $g(z)F(z)$ .

Let us illustrate our method with a small sized problem.

**Example 1**  $d = 1$ ,  $z = x$ .  $g(x) = \exp(-x + 1) \int_0^\infty \exp(xt - t^3) dt$ . The function  $g(x)$  satisfies the differential equation  $(3\partial_x^2 + 6\partial_x + (3 - x)) \bullet g = \exp(-x + 1)$ , which can be obtained by an integration algorithm for  $D$ -modules [9]. The

holonomic rank is 2 and we use a set of standard monomials  $S = \{1, \partial_x\}$  and we have

$$\frac{dG}{dx} = \begin{pmatrix} 0 & 1 \\ (-3+x)/3 & -2 \end{pmatrix} G + \begin{pmatrix} 0 \\ \exp(-x+1)/3 \end{pmatrix}$$

This system is obtained by the normal form algorithm in the ring  $R$  [13]. We note that it is easy to generalize our algorithm for a holonomic function which satisfies inhomogeneous holonomic system. Note that  $\frac{dg}{dx} = \nabla g = \begin{pmatrix} 0 & 1 \end{pmatrix} G$ . We evaluate  $G(0) = (g(0), g'(0))^T$  by a numerical integration method;  $\bar{G}(0) = (2.427, -1.20)^T$ . We apply the holonomic gradient descent in the search domain  $E = [0, 5]$  with  $h_k = -0.1$ ,  $H_k = 1$  and the 4th order Runge-Kutta method and obtain  $x = e = 3.4$  and  $g(e) = 1.016$  as the minimum in this domain.

The holonomic gradient descent is nothing but Euler's method when the number of variables is 1.

As we have seen, by utilizing integration algorithms, we can apply the holonomic gradient descent for a large class of optimization problems including integrals with parameters. However, integration algorithms require huge computational resources and we can solve only relatively small sized problems. Therefore, if we want to apply our method to larger problems for holonomic functions, we need to find systems of differential equations and Pfaffian equations without utilizing general algorithms. In fact, we will study a system of differential equations and Pfaffian equations for the Fisher-Bingham integral in the following sections to apply our method to a maximal likelihood estimate problem.

### 3 Fisher-Bingham Integral on $S^n$

We denote by  $S^n(r)$  the  $n$ -dimensional sphere with the radius  $r$  in the  $n+1$  dimensional Euclidean space. Let  $x$  be a  $(n+1) \times (n+1)$  symmetric matrix and  $y$  a row vector of length  $n+1$ . We are interested in the following integral with the parameters  $x, y, r$ .

$$F(x, y, r) = \int_{S^n(r)} \exp(t^T x t + y t) |dt| \quad (5)$$

Here,  $t$  is the column vector  $(t_1, \dots, t_{n+1})^T$  and  $|dt|$  is the standard measure on the sphere. For example, in case of  $n = 1$ , the measure  $|dt|$  is  $r d\theta$  in the polar coordinate system  $t_1 = r \cos \theta, t_2 = r \sin \theta$ . We call the integral (5) *the Fisher-Bingham integral* on the sphere  $S^n(r)$ .

We denote by  $x_{ii}$  the  $i$ -th diagonal entry of the matrix  $x$  and by  $x_{ij}/2$  the  $(i, j)$ -th entry (or  $(j, i)$ -th entry) of the matrix  $x$ . Then, we can regard the function (the Fisher-Bingham integral)  $F(x, y, r)$  as the function of  $x_{ij}$  ( $1 \leq i \leq j \leq n+1$ ) and  $y_i$  ( $1 \leq i \leq n+1$ ) and  $r$ .

**Theorem 1** *The Fisher-Bingham integral  $F(x, y, r)$  is a holonomic function.*

*Proof.* We will prove it for  $n = 1$  to avoid complicated indices. The cases for  $n > 1$  can be shown analogously.

Put  $x_1 = r \cos \theta, x_2 = r \sin \theta$  (the polar coordinate system). Then, the invariant measure  $|dt|$  is written as  $rd\theta$ . Therefore,  $F(x, y, r) = \int_0^{2\pi} e^{g(x, y, r, \theta)} r d\theta$  where  $g(x, y, r, \theta) = x_{11}r^2 \cos^2 \theta + x_{12}r^2 \cos \theta \sin \theta + x_{22}r^2 \sin^2 \theta + y_1 r \cos \theta + y_2 r \sin \theta$ . If we put  $s = \tan \frac{\theta}{2}$ , then  $\sin \theta = 2s/(s^2+1)$  and  $\cos \theta = (1-s^2)/(s^2+1)$  and  $d\theta = \frac{2}{1+s^2} ds$  (rational representation of trigonometric functions). Then, the integral  $F(x, y, r)$  can be written as

$$\int_{-\infty}^{\infty} h(x, y, r, s) ds, \quad h = e^{\tilde{g}(x, y, r, s)} \frac{2}{1+s^2}$$

where  $\tilde{g}$  is a rational function in  $x, y, r, s$ . It is known that the exponential of a rational function is a holonomic function and the product of holonomic functions is a holonomic function, then the integrand is a holonomic function in  $x, y, r, s$  (see, e.g., [11] and [12]). By Lemma 4 in the Appendix, there exists a differential operator  $\ell(x, y, r, \partial_{x_{ij}}) - \partial_s \ell_1(x, y, r, \partial_{x_{ij}}, \partial_s)$  depending only on  $x, \partial_{x_{ij}}, y, r, \partial_s$  which annihilates the integrand  $h$ . Therefore, we have  $\ell \bullet F(x, y, r) = [\ell_1 \bullet h]_{-\infty}^{\infty}$ . Since we can show that  $\partial_{x_{ij}}^m \partial_s^n \bullet h$  is a finite holonomic function at  $s = \pm\infty$  for any non-negative integers  $m$  and  $n$ , the function  $F(x, y, r)$  is annihilated by an ordinary differential operator of  $\partial_{x_{ij}}$  with parameters  $x, y, r$ . The existence of annihilating ordinary differential operators with respect to  $\partial_{y_i}$  and  $\partial_r$  can be shown analogously. This existence implies that  $F(x, y, r)$  is a holonomic function (see, e.g., [18, Theorem 2.4]). Q.E.D.

## 4 Holonomic system for the Fisher-Bingham Integral

In Example 1, we obtained a differential equation for the definite integral with parameters by a D-module algorithm. This algorithm works for any definite integral with a holonomic integrand, however, it requires huge computational resources. For the Fisher-Bingham integral, we can obtain a holonomic system of differential equations for the case of  $n = 1$  by our computer program. The case of  $n = 2$  is not feasible by our program. We obtain the following result for general  $n$  by utilizing an invariance of the Fisher-Bingham integral.

**Theorem 2** *The function  $F(x, y, r)$  is annihilated by the following system of linear partial differential operators.*

$$\partial_{x_{ij}} - \partial_{y_i} \partial_{y_j}, \quad (i \leq j) \tag{6}$$

$$\sum_{i=1}^{n+1} \partial_{x_{ii}} - r^2, \tag{7}$$

$$\begin{aligned} & x_{ij} \partial_{x_{ii}} + 2(x_{jj} - x_{ii}) \partial_{x_{ij}} - x_{ij} \partial_{x_{jj}} + \sum_{k \neq i, j} (x_{jk} \partial_{x_{ik}} - x_{ik} \partial_{x_{jk}}) \\ & + y_j \partial_{y_i} - y_i \partial_{y_j}, \quad (i < j, x_{kl} = x_{\ell k}), \end{aligned} \tag{8}$$

$$r\partial_r - 2\sum_{i \leq j} x_{ij}\partial_{x_{ij}} - \sum_i y_i\partial_{y_i} - n. \quad (9)$$

We note that operators of the form (6) can be written as

$$\partial^u - \partial^v, \quad Au = Av, \quad u, v \in \mathbf{N}^{(n+1)(n/2+2)}.$$

Here,  $A$  is the support matrix of the polynomial  $t^T xt + yt$  with respect to  $t$ . For example, in case of  $n = 1$ , the polynomial is  $x_{11}t_1^2 + x_{12}t_1t_2 + x_{22}t_2^2 + y_1t_1 + y_2t_2$  and the matrix  $A$  is

$$A = \begin{pmatrix} 2 & 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 1 \end{pmatrix}$$

of which column vectors stand for supports of the polynomial respectively.

*Proof.* Denote by  $g(x, y, t) = \exp(t^T xt + yt)$  the integrand of (5). The operator  $\partial_{x_{ij}} - \partial_{y_i}\partial_{y_j}$  annihilates  $g(x, y, t)$  because  $(\partial_{x_{ij}} - \partial_{y_i}\partial_{y_j}) \bullet g = (t_i t_j - t_i t_j)g = 0$ . On the sphere  $S^n(r)$ , we have an identity  $\sum_{i=1}^{n+1} t_i^2 = r^2$ . Hence  $\sum_{i=1}^{n+1} \partial_{x_{ii}} - r^2$  annihilates  $g(x, y, t)$  for  $t \in S^n(r)$ .

Let us prove (8). By the invariance of the measure  $|dt|$  with respect to the orthogonal group, we have  $F(PxP^T, yP^T, r) = F(x, y, r)$  for any orthogonal transformation  $P$  on  $S^n(r)$ . Let  $I_{n+1}$  be the  $(n+1) \times (n+1)$  identity matrix and  $e_{ij}$  be an  $(n+1) \times (n+1)$  matrix whose  $(k, l)$ -th entry  $(e_{ij})_{kl}$  is 1 if  $(i, j) = (k, l)$  and 0 else. Put  $P = \begin{pmatrix} \cos \epsilon & -\sin \epsilon \\ \sin \epsilon & \cos \epsilon \end{pmatrix} \oplus I_{n-1}$ . This is an  $(n+1) \times (n+1)$  orthogonal matrix and we have  $P = I_{n+1} + \epsilon(e_{12} - e_{21}) + O(\epsilon^2)$ . Hence we have

$$\begin{aligned} PxP^T &= (I + \epsilon(e_{12} - e_{21}))x(I + \epsilon(e_{21} - e_{12})) + O(\epsilon^2) \\ &= x + \epsilon(e_{12}x - e_{21}x + xe_{21} - xe_{12}) + O(\epsilon^2) \\ &= x + \epsilon \sum_{i \leq j} f_{ij}(x)(e_{ij} + e_{ji})/2 + O(\epsilon^2), \end{aligned}$$

where

$$f_{ij}(x) = \begin{cases} x_{12} & \text{if } i = j = 1, \\ 2(x_{22} - x_{11}) & \text{if } i = 1, j = 2, \\ -x_{12} & \text{if } i = j = 2, \\ x_{2j} & \text{if } i = 1, j \geq 3, \\ -x_{1j} & \text{if } i = 2, j \geq 3, \\ 0 & \text{if } j \geq i \geq 3, \end{cases}$$

and

$$yP^T = y + \epsilon \begin{pmatrix} y_2 & -y_1 & 0 \end{pmatrix} + O(\epsilon^2).$$

Differentiating the identity  $F(PxP^T, yP^T, r) - F(x, y, r) = 0$  by  $\epsilon$ , we obtain

$$0 = \left( \sum_{i \leq j} f_{ij}(x) \partial_{x_{ij}} + y_2 \partial_{y_1} - y_1 \partial_{y_2} \right) \bullet F + O(\epsilon).$$

Taking the limit  $\epsilon \rightarrow 0$ , we have (8) with  $i = 1$  and  $j = 2$ . By symmetry we have (8) for any  $i < j$ .



Finally we differentiate the identity  $\rho^n F(\rho^2 x, \rho y, r) = F(x, y, \rho r)$  by  $\rho$  and take the limit  $\rho \rightarrow 1$ . Then, we obtain

$$\left( n + 2 \sum_{i \leq j} x_{ij} \partial_{x_{ij}} + \sum_i y_i \partial_{y_i} \right) \bullet F = r \partial_r \bullet F$$

This shows that  $F$  is annihilated by (9). Q.E.D.

**Example 2** When  $n = 1$ , the system is written as follows.

$$\begin{aligned} & \partial_{x_{11}} - \partial_{y_1}^2, \partial_{x_{12}} - \partial_{y_1} \partial_{y_2}, \partial_{x_{22}} - \partial_{y_2}^2, \\ & \partial_{x_{11}} + \partial_{x_{22}} - r^2, \\ & x_{12} \partial_{x_{11}} + 2(x_{22} - x_{11}) \partial_{x_{12}} - x_{12} \partial_{x_{22}} + y_2 \partial_{y_1} - y_1 \partial_{y_2}, \\ & r \partial_r - 2(x_{11} \partial_{x_{11}} + x_{12} \partial_{x_{12}} + x_{22} \partial_{x_{22}}) - (y_1 \partial_{y_1} + y_2 \partial_{y_2}) - 1. \end{aligned}$$

**Example 3** When  $n = 2$ , the system is written as follows.

$$\begin{aligned} & \partial_{x_{11}} - \partial_{y_1}^2, \partial_{x_{12}} - \partial_{y_1} \partial_{y_2}, \partial_{x_{13}} - \partial_{y_1} \partial_{y_3}, \\ & \partial_{x_{22}} - \partial_{y_2}^2, \partial_{x_{23}} - \partial_{y_2} \partial_{y_3}, \partial_{x_{33}} - \partial_{y_3}^2, \\ & \partial_{x_{11}} + \partial_{x_{22}} + \partial_{x_{33}} - r^2, \\ & x_{12} \partial_{x_{11}} + 2(x_{22} - x_{11}) \partial_{x_{12}} - x_{12} \partial_{x_{22}} + x_{23} \partial_{x_{13}} - x_{13} \partial_{x_{23}} + y_2 \partial_{y_1} - y_1 \partial_{y_2}, \\ & x_{13} \partial_{x_{11}} + 2(x_{33} - x_{11}) \partial_{x_{13}} - x_{13} \partial_{x_{33}} + x_{23} \partial_{x_{12}} - x_{12} \partial_{x_{23}} + y_3 \partial_{y_1} - y_1 \partial_{y_3}, \\ & x_{23} \partial_{x_{22}} + 2(x_{33} - x_{22}) \partial_{x_{23}} - x_{23} \partial_{x_{33}} + x_{13} \partial_{x_{12}} - x_{12} \partial_{x_{13}} + y_3 \partial_{y_2} - y_2 \partial_{y_3}, \\ & r \partial_r - 2(x_{11} \partial_{x_{11}} + x_{12} \partial_{x_{12}} + x_{13} \partial_{x_{13}} + x_{22} \partial_{x_{22}} + x_{23} \partial_{x_{23}} + x_{33} \partial_{x_{33}}) \\ & \quad - (y_1 \partial_{y_1} + y_2 \partial_{y_2} + y_3 \partial_{y_3}) - 2. \end{aligned}$$

**Proposition 1** 1. The operators given in Theorem 2 generate a holonomic ideal in case of  $n = 1$  and  $n = 2$ .

2. The holonomic rank of the system for  $n = 1$  is 4. A set of standard monomials in  $R$  is

$$1, \partial_{y_1}, \partial_{y_2}, \partial_r.$$

3. The holonomic rank of the system for  $n = 2$  is 6. A set of standard monomials in  $R$  is

$$1, \partial_r, \partial_{y_3}, \partial_{y_2}, \partial_{y_1}, \partial_{x_{33}}.$$

The proposition can be shown by a calculation on a computer with applying algorithms for holonomic systems [14], [20, [toc.html](#)], [15].

We conjecture that the system of operators given in Theorem 2 generates a holonomic ideal in  $D$ , which is the ring of differential operators with polynomial coefficients. We can prove weaker result that they generate a zero dimensional ideal in  $R$ , which is sufficient for applying the holonomic gradient. This result

can also be used to derive Pfaffian equations. We will prove the zero dimensionality in the sequel.

For the Fisher-Bingham integral  $F(x, y, r)$ , let  $X = \{x, y, r\}$  be the set of all variables and  $\partial_X$  be the corresponding differential operators. Consider a ring  $R = \mathbf{C}(X)\langle\partial_X\rangle$ . Let  $I \subset R$  be the ideal generated by the operators (6) – (9) annihilating  $F(x, y, r)$  (Theorem 2). We show that the ideal  $I$  is zero-dimensional, that is, the quotient space  $R/I$  is a finite-dimensional vector space over  $\mathbf{C}(X)$ .

We denote  $\partial_{ij} = \partial_{x_{ij}}$  and  $\partial_i = \partial_{y_i}$  for simplicity. The symbol  $\partial_r$  is reserved for  $\partial/\partial r$ . It is easy to see that  $I$  is generated by

$$A_{ij} = \partial_{ij} - \partial_i \partial_j, \quad (10)$$

$$B = \sum_i \partial_i^2 - r^2, \quad (11)$$

$$C_{ij} = 2(x_{jj} - x_{ii})\partial_i \partial_j + x_{ij}\partial_i^2 - x_{ij}\partial_j^2 + \sum_{k \neq i, j} (x_{jk}\partial_i \partial_k - x_{ik}\partial_j \partial_k) + y_j \partial_i - y_i \partial_j, \quad (12)$$

$$E = r\partial_r - 2 \sum_{i \leq j} x_{ij} \partial_i \partial_j - \sum_i y_i \partial_i - n. \quad (13)$$

We write  $\ell_1 \equiv \ell_2$  if  $\ell_1 - \ell_2 \in I$ .

**Theorem 3** *Put  $S = \{1, \partial_1, \dots, \partial_{n+1}, \partial_1^2, \dots, \partial_n^2\}$  and let  $L_S$  be the vector space over  $\mathbf{C}(X)$  spanned by  $S$ . Then we have  $R = L_S + I$ . In particular, the ideal  $I$  is zero-dimensional.*

We prepare two lemmas. The proof is given later.

**Lemma 2** *For any  $i$  and  $j$ , we have  $\partial_i \partial_j \in L_S + I$ .*

**Lemma 3** *For any  $i, j$  and  $k$ , we have  $\partial_i \partial_j \partial_k \in L_S + I$ .*

We give a proof of Theorem 3 by using the lemmas. The proof implicitly uses a lexicographic order  $\prec$  such that  $\partial_k \prec \partial_{ij}$  and  $\partial_k \prec \partial_r$  for any  $k, i, j$ .

*Proof of Theorem 3.* We first show that  $R = \mathbf{C}(X)\langle\partial_1, \dots, \partial_{n+1}\rangle + I$ . Let  $f$  be an element of  $R$ . If a term of  $f$  is written as  $g\partial_{ij}$  with  $g \in R$ , then we can replace  $g\partial_{ij}$  with  $g\partial_i \partial_j$  because  $\partial_{ij} \equiv \partial_i \partial_j$ . By induction, there exists some  $f' \in R$  without  $\partial_{ij}$  such that  $f \equiv f'$ . If  $f'$  contains  $\partial_r$ , we can replace  $\partial_r$  with a polynomial of  $\{\partial_k\}$  by the annihilator (13). By induction, there exists some  $f'' \in \mathbf{C}(X)\langle\partial_1, \dots, \partial_{n+1}\rangle$  such that  $f \equiv f' \equiv f''$ . This proves  $R = \mathbf{C}(X)\langle\partial_1, \dots, \partial_{n+1}\rangle + I$ . Now we show that  $\mathbf{C}(X)\langle\partial_1, \dots, \partial_{n+1}\rangle + I = L_S + I$ . Let  $f = \prod_{i=1}^{n+1} \partial_i^{\beta_i}$  be any monomial in  $\mathbf{C}(X)\langle\partial_1, \dots, \partial_{n+1}\rangle$  with the total degree  $|\beta| = \sum_{i=1}^{n+1} \beta_i$ . If  $|\beta| \leq 1$ , clearly  $f \in L_S \subset L_S + I$ . If  $|\beta| = 2$ , Lemma 2 shows  $f \in L_S + I$ . If  $|\beta| \geq 3$ , then by Lemma 3 there is  $f'$  with the total degree less than or equal to  $|\beta| - 1$  such that  $f \equiv f'$ . By induction, we have some  $f'$  with the total degree less than or equal to 2 such that  $f \equiv f' \in L_S + I$ . This proves Theorem 3. Q.E.D.

Now we prove Lemma 2 and Lemma 3.

*Proof of Lemma 2.* From the definition of  $S$ , it is obvious that  $\partial_i^2 \in L_S$  for  $1 \leq i \leq n$ . Since  $\partial_{n+1}^2 \equiv -\sum_{i=1}^n \partial_i^2 + r$  by (11), we have  $\partial_{n+1}^2 \in L_S + I$ . Now we prove that  $\partial_i \partial_j \in L_S + I$  for any  $1 \leq i < j \leq n+1$ . We use the annihilator  $C_{ij}$  in (12). Denote the quadratic part of  $C_{ij}$  by  $\sum_{k < l} P_{ij,kl} \partial_k \partial_l$ , where  $P_{ij,kl} = P_{ij,kl}(x, y, r) \in \mathbf{C}(X)$ . Since 1 and  $\partial_k$  are in  $L_S + I$ , we have

$$\sum_{k < l} P_{ij,kl} \partial_k \partial_l \in L_S + I.$$

To show  $\partial_i \partial_j \in L_S + I$ , it is sufficient to prove that the determinant of the coefficient matrix  $(P_{ij,kl})_{i < j, k < l}$  is a non-zero element in  $\mathbf{C}(X)$ . We evaluate  $P_{ij,kl}$  at a point  $(x, y, r) = (\bar{x}, \bar{y}, \bar{r})$  such that  $\bar{x}_{ii} \neq \bar{x}_{jj}$  and  $\bar{x}_{ij} = 0$  for any  $i < j$ . Then we obtain

$$P_{ij,kl}(\bar{x}, \bar{y}, \bar{r}) = \begin{cases} 2(\bar{x}_{jj} - \bar{x}_{ii}) & \text{if } (i, j) = (k, l), \\ 0 & \text{else.} \end{cases}$$

In particular,  $P_{ij,kl}(\bar{x}, \bar{y}, \bar{r})$  is a diagonal matrix and its determinant is  $\prod_{i < j} 2(\bar{x}_{jj} - \bar{x}_{ii}) \neq 0$ . Hence the determinant of  $(P_{ij,kl})$  is non-zero in  $\mathbf{C}(X)$ . Q.E.D.

*Proof of Lemma 3.* Consider an operator  $\partial_i \partial_j \partial_k$  with  $i \leq j \leq k$ . If  $j = k = n+1$ , then  $\partial_i \partial_{n+1}^2 \equiv \partial_i(-\sum_{l=1}^n \partial_l^2 + r^2)$ . Hence we can assume  $j \leq n$ . By using the operator  $C_{ij}$  in (12), we define an operator  $G_{ijk}$  by

$$G_{ijk} = \begin{cases} \partial_i C_{jk} & \text{if } j < k, \\ \partial_j C_{ij} & \text{if } i < j = k(\leq n), \\ \partial_{n+1} C_{i,n+1} & \text{if } i = j = k(\leq n) \end{cases}$$

Then  $G_{ijk} \equiv 0$ . As in the proof of Lemma 2, denote the cubic term of  $G_{ijk}$  by  $\sum_{a \leq b \leq c; b \leq n} P_{ijk,abc} \partial_a \partial_b \partial_c$ . Since all quadratic terms are in  $L_S + I$ , we obtain

$$\sum_{a \leq b \leq c; b \leq n} P_{ijk,abc} \partial_a \partial_b \partial_c \in L_S + I.$$

It is sufficient to show that  $\det(P_{ijk,abc})$  is a non-zero element in  $\mathbf{C}(X)$ . As in the proof of Lemma 2, we evaluate  $P_{ijk,abc}$  at a point  $(\bar{x}, \bar{y}, \bar{r})$  such that  $\bar{x}_{ii} \neq \bar{x}_{jj}$  and  $\bar{x}_{ij} = 0$  for any  $i < j$ . Then, with a little effort, we obtain

$$P_{ijk,abc}(\bar{x}, \bar{y}, \bar{r}) = \begin{cases} 2(\bar{x}_{kk} - \bar{x}_{jj}) \delta_{ia} \delta_{jb} \delta_{kc} & \text{if } j < k, \\ 2(\bar{x}_{jj} - \bar{x}_{ii}) \delta_{ia} \delta_{jb} \delta_{jc} & \text{if } i < j = k(\leq n), \\ -2(\bar{x}_{n+1,n+1} - \bar{x}_{ii}) \{ \delta_{ia} \delta_{ib} \delta_{ic} \\ + \sum_{h < i} \delta_{ha} \delta_{hb} \delta_{ic} + \sum_{i < h \leq n} \delta_{ia} \delta_{hb} \delta_{hc} \} & \text{if } i = j = k(\leq n). \end{cases}$$

Remark that all the diagonal elements  $P_{ijk,ijk}$  are non-zero. We sort indices  $\{(i, j, k) \mid i \leq j \leq k, j \leq n\}$  in such a way that  $(i, i, i)$  is greater than  $(j, k, l)$  unless  $j = k = l$ . Then we can conclude that  $P_{ijk,abc}(\bar{x}, \bar{y}, \bar{r}) = 0$  if  $(i, j, k)$  is less than  $(a, b, c)$ . Hence  $P_{ijk,abc}(\bar{x}, \bar{y}, \bar{r})$  is a triangular matrix and its determinant is product of the diagonal elements. This proves that  $\det(P_{ijk,abc})$  is a non-zero element in  $\mathbf{C}(X)$ . Q.E.D.

## 5 Computational Results

Let us apply the holonomic gradient descent to minimize the holonomic function

$$F(x, y, 1) \exp \left( - \sum_{1 \leq i \leq j \leq n} S_{ij} x_{ij} - \sum_i S_i y_i \right) \quad (14)$$

with respect to  $x$  and  $y$  for given data  $((S_{ij})_{i \leq j}, (S_i))$ . Here  $F(x, y, 1)$  is the Fisher-Bingham integral (5) with  $r = 1$ .

First we describe the background in statistics. This paragraph can be skipped for the reader interested only in computational results. *The Fisher-Bingham family* on the sphere  $S^n(1)$  is defined by the set of probability density functions

$$p(t|x, y) = F(x, y, 1)^{-1} \exp(t^\top x t + y t) \quad (15)$$

with respect to the standard measure  $|dt|$  on  $S^n(1)$ . Since  $\int_{S^n(1)} p(t|x, y) |dt| = 1$ , the function  $p(t|x, y)$  is actually a probability density function. We note that the parameter  $x$  has redundancy. In fact, for any real number  $c$  the density function  $p(t|x + cI, y)$  is equal to  $p(t|x, y)$ , where  $I$  denotes the identity matrix. A *sample* refers to a set of points  $\{t(1), \dots, t(N)\}$  on  $S^n(1)$ , where  $N \geq 1$  is called the sample size. Assume that the sample is distributed according to  $\prod_{\nu=1}^N p(t(\nu)|x, y)$  (independently identically distributed). To estimate the unknown parameter  $(x, y)$  from the sample is a main problem in statistics. An established method is *the maximum likelihood method (MLE)* that maximizes a function  $\prod_{\nu=1}^N p(t(\nu)|x, y)$  with respect to  $(x, y)$ . The MLE is equivalent to minimize the function (14) with  $S_{ij} = N^{-1} \sum_{\nu=1}^N t_i(\nu) t_j(\nu)$  and  $S_i = N^{-1} \sum_{\nu=1}^N t_i(\nu)$ . It is known that the logarithm of (14) is convex (see e.g. [2]) and therefore a local minimum at an interior point is actually the global minimum. Although gradient systems on probability families for optimization are considered by [8], difficulty of computing the integral  $F$  is not taken into account. See [7] for details on the Fisher-Bingham family and other probability families on the sphere. We test two examples, astronomical data and magnetism data. The astronomical data consist of the locations of 188 stars of magnitude brighter than or equal to 3.0. The data is available from the Bright Star Catalog (5th Revised Ed.) distributed from the Astronomical Data Center. The magnetism data is analyzed in [3] and [6].

The data and programs to test the following examples can be obtained from [20].

**Remark 1** Let  $e_i$  be the  $i$ -th standard vector. We note that  $G(z^{(k)} + e_i h_k)$  can approximately be obtained by evaluating  $P_i(z^{(k)})G(z^{(k)})h_k$ . In our implementation in [20], we choose a search direction  $d^{(k)}$  which is parallel to a coordinate axis. In other words, if the direction  $h_k e_i$  is chosen, then we move to the direction as long as  $g$  decreases to the direction  $h_k e_i$ . Because  $P_i$  is a matrix of a huge size and the computational cost of restricting the variables  $z_j$ ,  $j \neq i$  in  $P_i$

to numbers is extremely high in the problem of Fisher-Bigham integral and our implementation.

**Astronomical data:** We consider the problem to minimize

$$F(x, y, 1) \exp \left( - \sum_{1 \leq i \leq j \leq 3} S_{ij} x_{ij} - \sum_i S_i y_i \right)$$

on

$$\begin{aligned} & (x_{11}, x_{12}, x_{13}, x_{22}, x_{23}, x_{33}, y_1, y_2, y_3) \\ \in & E = [-30, 10] \times [-30, 10] \times [-30, 10] \times [-30, 10] \times [-30, 20] \times [-30, -0.01] \\ & \times [-30, -0.01] \times [-30, -0.001] \times [-30, 10] \end{aligned}$$

where

$$\begin{aligned} & (S_{11}, S_{12}, S_{13}, S_{22}, S_{23}, S_{33}, S_1, S_2, S_3) \\ = & (0.3119, 0.0292, 0.0707, 0.3605, 0.0462, 0.3276, -0.0063, -0.0054, -0.0762). \end{aligned}$$

The result is that the minimum 11.68573121328159669 is taken at

$$x = \begin{pmatrix} -0.161 & 0.3377/2 & 1.1104/2 \\ 0.3377/2 & 0.2538 & 0.6424/2 \\ 1.1104/2 & 0.6424/2 & -0.0928 \end{pmatrix}, \quad y = (\underline{-0.019}, \underline{-0.0162}, -0.2286) \text{ with}$$

the grid size 0.05 and the 4th order Runge-Kutta method for solving the Pfaffian system numerically (see Fig. 1), where the values near the border are underlined. A starting point is found by a quadratic approximation of  $F(x, y, 1)$ , which is exactly calculated from the moments of the uniform distribution on the sphere, and solving the optimization problem for the quadratic polynomial.

We briefly discuss the statistical meaning of the result. The spectral decomposition of  $x$  is  $x = \sum_{i=1}^3 \lambda_i z_i z_i^T$  with

$$(\lambda_1, \lambda_2, \lambda_3) = (0.7047, -0.0103, -0.6944)$$

and

$$(z_1, z_2, z_3) = \begin{pmatrix} -0.5063 & 0.5055 & 0.6987 \\ -0.6181 & -0.7777 & 0.1148 \\ -0.6014 & 0.3737 & -0.7061 \end{pmatrix}.$$

From the decomposition the density function (15) is high around  $\pm z_1$  and low around  $\pm z_3$ . The effect of  $y$  is small because  $|y| = 0.230$  is smaller than  $|\lambda_i|$ 's.

As we have seen, we have determined the model parameters  $x$  and  $y$  by the holonomic gradient descent successfully. However, the computation poses us two future problems to make the method stronger and more useful. The first problem is to determine the search domain  $E$  of  $x$  and  $y$  automatically. We set the search domain in this case by a help of human intuition and numerical evaluations of the target function at several points. The second problem is to move over the singular locus of the Pfaffian system without numerical instability. In this case, we pose the conditions  $x_{33} \leq -0.01$ ,  $y_1 \leq -0.01$  and  $y_2 \leq -0.001$ , because the variety  $x_{33} = y_1 = y_2 = 0$  lies in the singular locus of the Pfaffian system.

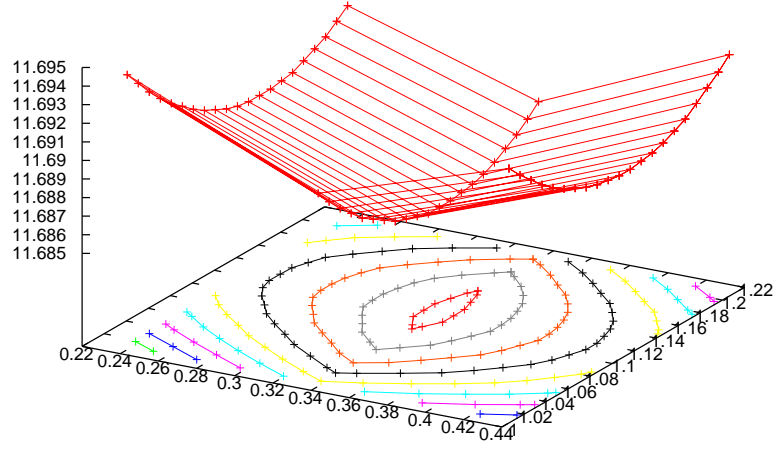


Figure 1: Graph of the target function with varying  $x_{12}$  and  $x_{13}$  around the minimal point for astronomical data.

### Magnetism data

We consider the problem to minimize

$$F(x, y, 1) \exp \left( - \sum_{1 \leq i \leq j \leq 3} S_{ij} x_{ij} - \sum_i S_i y_i \right)$$

on

$$\begin{aligned} & (x_{11}, x_{12}, x_{13}, x_{22}, x_{23}, x_{33}, y_1, y_2, y_3) \\ \in & E = [-30, 30] \times [-30, 30] \times [-30, 30] \times [-30, 30] \times [-30, 30] \times [-30, -0.01] \\ & \times [-30, 30] \times [-32, -0.001] \times [-30, 32] \end{aligned}$$

where

$$\begin{aligned} & (S_{11}, S_{12}, S_{13}, S_{22}, S_{23}, S_{33}, S_1, S_2, S_3) \\ = & (0.045, -0.075, 0.014, 0.921, -0.122, 0.034, 0.082, -0.959, 0.131). \end{aligned}$$

The result is that the minimum 0.4373096253840751950 is taken at

$$x = x_o = \begin{pmatrix} 7.065 & -0.032/2 & 3.422/2 \\ -0.032/2 & 5.339 & 24.922/2 \\ 3.422/2 & 24.922/2 & -13.693 \end{pmatrix}, \quad y = (1.642, \underline{-31.99}, \underline{31.992})$$

with the grid size 0.01 and the 4th order Runge-Kutta method. Although  $y_2$  and  $y_3$  are on the border with this grid size, we can observe that the change of the target value is relatively small, when we enlarge the domain. In fact, we started the holonomic gradient descent from the optimal point, obtained by Wood's method [19], [20, [toc.html](#)], which is

$$x = \begin{pmatrix} 5.985 & 8.478/2 & 2.902/2 \\ 8.478/2 & 6.869 & 16.732/2 \\ 2.902/2 & 16.732/2 & -12.853 \end{pmatrix}, y = (9.762, -28.770, 24.142). \text{ The op-}$$

timal value of the target function is 0.4421940620633763292. If we restart the holonomic gradient descent from the point  $x_o$  by recalculating the integral values, we get a new optimal point and the target value changes only about  $10^{-5}$ . Since the significant figures of the given data  $S_{ij}, S_i$  are 2 digits, we may conclude that there seems to be a variety which gives the optimal value of the target function. Our method finds a point in the variety and moves in the variety.

The statistical problems considered in this section can be solved by a different method. A. T. A. Wood [19] expressed the Fisher-Bingham integral of the case  $n = 2$  as a single integral with the integrand expressed by a modified Bessel function. He gives a method to solve a minimization problem equivalent to our problem (14) based on this single integral representation. We implement his method by the statistical computing system R and obtain analogous computational results with us. The program is obtainable from [20, [toc.html](#)].

Although our two statistical problems can be solved by his different method, the advantage of our approach is that our method is a general algorithm which can be applied to a broad class of problems, which will be presented in forthcoming papers, and is based on a holonomic system of differential equations. We note that this point of view of holonomic system has been emphasized by few literatures in statistics.

*Acknowledgements.* We thank to Prof. K.Takeda for comments on optimization methods.

## 6 Appendix: Introduction to Holonomic Ideals

Although we want to suppose people with different disciplines as readers of this paper, the theory and algorithms for holonomic ideals are not very popular and facts needed for the holonomic gradient descent are in diverse literatures. We will present an introductory overview on these well-known facts of holonomic ideals and algorithms (see [15] and its references for proofs and original articles).

We denote by  $D$  the ring of differential operators with polynomial coefficients

$$D = \mathbf{C}\langle x_1, \dots, x_d, \partial_1, \dots, \partial_d \rangle,$$

which is also called the Weyl algebra. This is an associative non-commutative ring and  $x_i$  and  $\partial_j$  have the commuting relations

$$x_i x_j = x_j x_i, \partial_i \partial_j = \partial_j \partial_i, \partial_i x_j = x_j \partial_i + \delta_{ij}$$

where  $\delta_{ij}$  is Kronecker's delta. Elements in  $D$  are often expressed by using the multi-index notation such as  $x^\alpha \partial^\beta = \prod_{i=1}^d x_i^{\alpha_i} \prod_{i=1}^d \partial_i^{\beta_i}$ .  $|\alpha|$  is defined by  $\alpha_1 + \dots + \alpha_d$ . By utilizing the commuting relations, any element of  $D$  can be transformed into the normally ordered form  $\sum_{(\alpha, \beta) \in E} c_{\alpha\beta} x^\alpha \partial^\beta$ . For example, the normally ordered form of  $\partial_1 x_1 \partial_1$  is  $x_1 \partial_1^2 + \partial_1$ . Elements of  $D$  acts on a function  $f(x_1, \dots, x_d)$  by

$$x^\alpha \partial^\beta \bullet f = x^\alpha \frac{\partial^{|\beta|} f}{\partial x_1^{\beta_1} \dots \partial x_d^{\beta_d}}$$

where we denote by  $\bullet$  the action.

Let us introduce one more important ring  $R$ , which we call the ring of differential operators with rational function coefficients,

$$R = \mathbf{C}(x_1, \dots, x_d) \langle \partial_1, \dots, \partial_d \rangle$$

where we denote by  $\mathbf{C}(x_1, \dots, x_d)$  the field of rational functions in  $x_1, \dots, x_d$ . This is also an associative non-commutative ring and the commuting relations are  $\partial_i \partial_j = \partial_j \partial_i$  and  $\partial_i a(x) = a(x) \partial_i + \frac{\partial a}{\partial x_i}$  for  $a(x) \in \mathbf{C}(x_1, \dots, x_d)$ .

The theory of Gröbner basis (see, e.g., [4]) can be easily generalized in  $D$  and  $R$  as long as orders satisfy some conditions. Since we do not need consider general orders, we fix the order to the graded reverse lexicographic order  $\prec$  among monomials  $\partial^\beta$  in the sequel. In case of  $d = 2$ , we have

$$1 \prec \partial_2 \prec \partial_1 \prec \partial_2^2 \prec \partial_1 \partial_2 \prec \partial_1^2 \prec \dots$$

Let us explain some facts about Gröbner bases in  $R$ , which are used in this paper. For  $f \in R$ , the leading term (the initial term) with respect to  $\prec$  is denoted by  $\text{in}_\prec(f)$  and we regard this element as an element in  $\mathbf{C}(x_1, \dots, x_d)[\xi_1, \dots, \xi_d]$  where  $\xi_i$  and  $x_j$  commute each other. For example, when  $f = (x_1 + x_2) \partial_1^2 \partial_2 + (x_2^4 + 1) \partial_2$ , we have  $\text{in}_\prec(f) = (x_1 + x_2) \xi_1^2 \xi_2$ . We say that  $a(x) \xi^\beta$  divides  $b(x) \xi^{\beta'}$  when  $\beta_i \leq \beta'_i$  for all  $i$ . We call the following algorithm *the normal form algorithm (the division algorithm)*.

**Algorithm 3** (NormalForm( $f, G$ ))

Input:  $f, G = \{g_1, \dots, g_m\}$

Output: The normal form  $r$  (remainder) and quotients  $q_1, \dots, q_m$ , which satisfy the following conditions (a)  $f = \sum_{i=1}^m q_i g_i + r$  in  $R$ , (b)  $f \succeq q_i g_i$ , (c)  $\text{in}_\prec(g_i)$  does not divide any term of  $r|_{\partial \rightarrow \xi}$  for all  $i$ .

1.  $r \leftarrow 0, q_i \leftarrow 0$ .
2. Call **wNormalForm**( $f, G$ ). We suppose that the output is  $r', q'_1, \dots, q'_m$ .
3.  $f \leftarrow r' - \text{in}_\prec(r'), r \leftarrow r + \text{in}_\prec(r'), q_i \leftarrow q_i + q'_i$ . If  $f = 0$ , then return  $r, q_1, \dots, q_m$  else goto 2.

**Algorithm 4** (wNormalForm( $f, G$ ))



1.  $r \leftarrow f, q_i \leftarrow 0$
2. If there exists  $i$  such that  $\text{in}_{\prec}(g_i)$  divides  $\text{in}_{\prec}(r)$  then  
 $r \leftarrow r - c(x)\partial^\beta g_i$  where  $c(x)\partial^\beta$  is chosen so that  $\text{in}_{\prec}(r) - c(x)\xi^\beta \text{in}_{\prec}(g_i) = 0$ ;  
 $q_i \leftarrow q_i + c(x)\partial^\beta$ ;  
else return  $r, q_1, \dots, q_m$ .
3. goto 2.

**Example 4** We compute the normal form of  $f = \partial_1 \partial_2^3$  by  $g_1 = \partial_1 \partial_2 + 1$ ,  $g_2 = \frac{\partial_1 \partial_2^2}{2x_2} - \partial_1 + 3\partial_2 + 2x_1$  with the graded reverse lexicographic order. Since we have

$$\begin{aligned} \partial_1 \partial_2^3 - \partial_2^2 g_1 &= -\partial_2^2 \\ -\partial_2^2 + \frac{1}{2x_2} g_2 &= \frac{1}{2x_2} (-\partial_1 + 3\partial_2 + 2x_1) =: f^*, \end{aligned}$$

the normal form is  $f^*$  and  $q_1 = \partial_2^2$  and  $q_2 = -\frac{1}{2x_1}$ . This example is taken from [11].

Let  $I$  be a left ideal in  $R$ . A finite set  $G = \{g_1, \dots, g_m\}$ ,  $g_i \in R$  is called a Gröbner basis of  $I$  with respect to  $\prec$  when  $\langle \text{in}_{\prec}(g_1), \dots, \text{in}_{\prec}(g_m) \rangle = \langle \text{in}_{\prec}(f) \mid f \in I \rangle$ . Here,  $\langle h_1, \dots, h_m \rangle$  is the set  $\sum_{i=1}^m \mathbf{C}(x_1, \dots, x_d)[\xi_1, \dots, \xi_d] h_i$ , which is the ideal generated by  $h_1, \dots, h_m$  in  $\mathbf{C}(x_1, \dots, x_d)[\xi_1, \dots, \xi_d]$ . A Gröbner basis can be obtained by the Buchberger algorithm. The proof is analogous with the case of the ring of polynomials (see, e.g., [4, Chapter 2]).

Let  $G$  be a Gröbner basis. The element  $\partial^\beta$  is called a standard monomial when none of  $\text{in}_{\prec}(g)$ ,  $g \in G$  divides  $\xi^\beta$ . Any normal form is a sum of standard monomials over  $\mathbf{C}(x_1, \dots, x_d)$ .

**Example 5** This is a continuation of the previous example. Put  $g_3 = \frac{\partial_1^2}{2} - 3\partial_1 \partial_2 - 2x_1 \partial_1 + 2x_2 \partial_2 - 2$ . Then, the set  $\{g_1, g_2, g_3\}$  is a Gröbner basis of the left ideal in  $R$  generated by  $g_1$  and  $g_2$ . The set of the standard monomials is  $\{1, \partial_1, \partial_2\}$ .

The output  $r$  of the normal form algorithm depends on which index  $i$  we choose in the step 2 in the algorithm `wNormalForm`.

**Theorem 4** Let  $f$  be an element of  $R$ . If  $G$  is a Gröbner basis, then the normal form  $r$  of  $f$  by  $G$  is unique.

*Proof.* Suppose that we have two different normal forms  $r_1$  and  $r_2$ . Since we have  $r_1 - r_2 \in I$ ,  $\text{in}_{\prec}(r_1 - r_2)$  is divisible by an  $\text{in}_{\prec}(g_i)$  by the definition of Gröbner basis. But it contradicts to that  $r_i$  is a sum of standard monomials over  $\mathbf{C}(x_1, \dots, x_d)$ . Q.E.D.

When the number of the standard monomials is finite, the ideal  $I$  is called a *zero-dimensional ideal*. It follows from Theorem 4 that the number is equal to the dimension of  $R/I$  as the vector space over  $\mathbf{C}(x_1, \dots, x_d)$  (see, e.g., [4,

Chapter 5]). It implies that the number of the standard monomials does not depend on Gröbner bases. The dimension is called the *holonomic rank* of  $I$ .

We call  $c(x)\partial^\beta$ ,  $0 \neq c(x) \in \mathbf{C}(x_1, \dots, x_d)$ , a non-monic standard monomial when  $\partial^\beta$  is a standard monomial. Let  $S = \{s_1 = 1, s_2, \dots, s_p\}$  be a set of (independent) non-monic standard monomials of the Gröbner basis  $G$  such that  $p = \sharp S = \dim_{\mathbf{C}(x_1, \dots, x_d)} R/RG$ . Put  $Q = (s_i \bullet g \mid s_i \in S)^T$ . In order to apply holonomic gradient descent, we need to compute the  $p \times p$  matrix  $P_i$  in the Pfaffian equations

$$\frac{\partial Q}{\partial x_i} = P_i Q, \quad i = 1, \dots, d.$$

which is (4) in the main text. To obtain the matrix  $P_i$ , we apply the normal form algorithm to  $\partial_i s_j$ . Then, the coefficient of the normal form of  $\partial_i s_j$  with respect to  $s_k$  is the  $(j, k)$ -th element of  $P_i$ . This is the step 2 of the Algorithm 1 in the main text.

**Example 6** This is a continuation of the previous example. We choose  $S = \{1, x_1 \partial_1, x_2 \partial_2\}$ . Then, we obtain

$$P_1 = \begin{pmatrix} 0 & \frac{1}{x} & 0 \\ -x & \frac{2x^2+1}{x} & -2x \\ -y & 0 & 0 \end{pmatrix}, P_2 = \begin{pmatrix} 0 & 0 & \frac{1}{y} \\ -x & 0 & 0 \\ -x & \frac{1}{x} & \frac{-1}{y} \end{pmatrix}$$

where  $x = x_1$  and  $y = x_2$ . We can utilize several packages to perform this computation. Among them, we use the package “yang” [13] on Risa/Asir<sup>1</sup>, because it can perform a large scale computation, which is required in our applications. The code to obtain the result above is

```
import("yang.rr");
def ex1() {
  yang.define_ring([x,y]);
  L1=dx*dy+1;
  L2=dx^2-2*x*dx+2*y*dy+1;
  L3=2*y*dy^2+3*dy-dx+2*x;
  L=[L1,L2,L3];
  L=yang.util_pd_to_euler(L,[x,y]);
  L=map(nm,L);
  L=map(dp_ptod,L,[dx,dy]);
  G=yang.buchberger(L);
  S1=yang.constant(1);
  Sx=yang.operator(x);
  Sy=yang.operator(y);
  Base=[S1,Sx,Sy];
  Pf=yang.pfaffian(Base,G);
  return Pf;
}
ex1();
```

Since we have  $\partial_1 = \frac{1}{x_1} s_2$  and  $\partial_2 = \frac{1}{x_2} s_2$ , the gradient  $\nabla g = \begin{pmatrix} \frac{\partial g}{\partial x} \\ \frac{\partial g}{\partial y} \end{pmatrix}$  is equal to  $AG$  where the matrix  $A = (a_{ij})$  is  $\begin{pmatrix} 0 & \frac{1}{x_1} & 0 \\ 0 & 0 & \frac{1}{x_2} \end{pmatrix}$ .

---

<sup>1</sup>[14], <http://www.math.kobe-u.ac.jp/Asir>

We call a function  $F$  a *holonomic function* when it satisfies ordinary differential equations for all variables. In other words,  $F$  satisfies

$$\sum_{k=0}^{r_i} a_k^i(x_1, \dots, x_d) \partial_i^k \bullet F = 0, \quad a_k^i \in \mathbf{C}[x_1, \dots, x_d], \quad i = 1, \dots, d. \quad (16)$$

The set of operators in  $R$  which annihilate a function  $F$  is a left ideal in  $R$ . In fact, if  $\ell_1 \bullet F = \ell_2 \bullet F = 0$ , then we have  $(\ell_1 + \ell_2) \bullet F = 0$ , and if  $\ell \bullet F = 0$ , then  $(h\ell) \bullet F = 0$  for all  $h \in R$ . We denote the set by  $\text{Ann}_R F$ . When the function  $F$  is holonomic,  $\text{Ann}_R F$  contains ordinary differential equations (16). Therefore, the number of standard monomials of a Gröbner basis of  $\text{Ann}_R F$  is less than or equal to  $\prod_{i=1}^d r_i$ . In other words, we have  $\dim_{\mathbf{C}(x_1, \dots, x_d)} R / \text{Ann}_R F \leq \prod_{i=1}^d r_i$ . Conversely, we have the following theorem.

**Theorem 5** *Let  $I$  be a left ideal in  $R$ . If  $m := \dim_{\mathbf{C}(x_1, \dots, x_d)} R/I$  is finite, then the left ideal  $I$  contains an ordinary differential operator for any variable  $x_i$ .*

*Proof.* 1,  $\partial_i, \partial_i^2, \dots, \partial_i^m$  are linearly dependent in  $R/I$ , which we regard as a vector space over  $\mathbf{C}(x_1, \dots, x_d)$ . This implies that there exist rational functions  $c_k(x)$  such that  $\sum_{k=0}^m c_k(x) \partial_i^k \in I$ . Q.E.D.

This theorem is an analogy of the elimination theorem. The elimination in  $R$  can be done by an analogous method in case of the ring of polynomials (see, e.g., [4, Chapter 3]).

We have worked in the ring  $R$ . If we need to consider integrals of  $F$ , we need the theory and algorithms for the Weyl algebra  $D$ . Let us proceed on a discussion on  $D$ .

We first note that we can easily generalize the Gröbner basis theory for term orders  $\prec$  in  $D$ . For example, in case of  $d = 2$ , the Gröbner basis theory works for the graded reverse lexicographic order such that  $1 \prec x_1 \prec x_2 \prec \partial_1 \prec \partial_2 \prec x_1^2 \prec \dots$ .

We introduce the notion of a holonomic ideal. Let  $F_k$  be the set of elements in  $D$  of which order is less than or equal to  $k$ . In other words,  $F_k$  is a  $\mathbf{C}$ -vector space spanned by  $x^\alpha \partial^\beta$ ,  $|\alpha| + |\beta| \leq k$ .  $\{F_k\}$  is called the Bernstein filtration. A left ideal  $I$  in  $D$  is called a *holonomic ideal* when  $\dim_{\mathbf{C}} F_k / F_k \cap I = O(k^d)$  for sufficiently large numbers  $k$ . The quotient  $D/I$  is called a *holonomic  $D$ -module* when  $I$  is a holonomic ideal. We note that the dimension agrees with the number of standard monomials of which total degree is less than or equal to  $k$  with respect to a Gröbner basis of  $I$  by the graded reverse lexicographic order (see, e.g., [4, Chapter 9]).

**Lemma 4** *Let  $I$  be a holonomic ideal in the ring of differential operators  $D = \mathbf{C}\langle x_1, \dots, x_d, \partial_1, \dots, \partial_d \rangle$ . We choose a set of  $d + 1$  variables from the set  $\{x_1, \dots, x_d, \partial_1, \dots, \partial_d\}$  and denote it by  $V$ . Then, the elimination ideal  $I \cap \mathbf{C}\langle V \rangle$  contains a non-zero element.*

*Proof.* Consider the  $\mathbf{C}$ -linear map

$$\rho_k : \mathbf{C}\langle V \rangle \cap F_k \ni \ell \mapsto [\ell] \in F_k / F_k \cap I$$

The dimension of the  $\mathbf{C}$ -vector space  $\mathbf{C}\langle V \rangle \cap F_k$  is  $\binom{d+1+k}{d+1} = O(k^{d+1})$ . On the other hand, we have  $\dim_{\mathbf{C}} F_k/F_k \cap I = O(k^d)$  because  $I$  is a holonomic ideal. Since  $\dim_{\mathbf{C}} \text{Im } \rho_k = \dim_{\mathbf{C}} \mathbf{C}\langle V \rangle \cap F_k - \dim_{\mathbf{C}} \text{Ker } \rho_k$ , we conclude that the vector space  $\text{Ker } \rho_k$  contains a non-zero element. Q.E.D.

When  $I$  is a holonomic ideal, the number of standard monomials is infinite in general. It is natural to ask if there is a zero-dimensional ideal in  $D$ . However, the following theorem claims that the holonomic ideals are the biggest ideals and there is no zero-dimensional ideal in  $D$ .

**Theorem 6** (Bernstein inequality) *Let  $I$  be a left ideal in  $D$ . Suppose that  $I \neq D$ . There exists a constant  $p$  such that  $\dim_{\mathbf{C}} F_k/F_k \cap I = O(k^p)$  for sufficiently large  $k$  and the inequality  $p \geq d$  holds.*

Let us explain a relation of a holonomic ideal in  $D$  and a zero dimensional ideal in  $R$ . For a left ideal  $I$  in  $D$ , we denote by  $RI$  the left ideal in  $R$  generated by elements in  $I$ . It follows from the Lemma 4 that if  $I$  is a holonomic ideal, then  $I$  contains an ordinary differential operator for any variable  $x_i$  and then  $RI$  is a zero-dimensional ideal. Conversely, we have the following theorem.

**Theorem 7** *If  $J$  is a zero-dimensional ideal in  $R$ , then  $J \cap D$  is a holonomic ideal in  $D$ .*

An elementary proof of this fact is found in the appendix of [18]. We emphasize that when we are given a set of generators of  $J$ , it is not necessarily a set of generators of  $J \cap D$ . The ideal  $J \cap D$  is called *the Weyl closure* of  $J$ . An algorithm to find a set of generators of the Weyl closure is given by H. Tsai (Algorithms for associated primes, Weyl closure, and local cohomology of  $D$ -modules. Lecture Notes in Pure and Appl. Math., 226, 169–194, Dekker, New York, 2002). Although we can make a lot of constructions for 0-dimensional ideals in  $R$ , for algorithms in  $D$  like  $D$ -module theoretic integration algorithms, we often require that *inputs are holonomic*. However, finding a set of generators of  $J \cap D$  requires a high complexity. It often makes computational bottlenecks.

**Example 7** We consider the function  $f(x, y, z) = \exp(1/g)$  where  $g = x^3 - y^2z^2$ . The function  $f$  is annihilated by first order operators

$$g^2\partial_x + 3x^2, g^2\partial_y - 2yz^2, g^2\partial_z - 2y^2z$$

The left ideal  $I$  generated by these operator is not holonomic. The Weyl closure  $J = RI \cap D$  is holonomic. The below is a Macaulay 2<sup>2</sup> script to check the holonomicity and find the Weyl closure of  $RI$ .

```
loadPackage "Dmodules"
D=QQ[x,y,z,dx,dy,dz, WeylAlgebra=>{x=>dx,y=>dy,z=>dz}];
I = ideal((x^3-y^2*z^2)^2*dx+3*x^2,
          (x^3-y^2*z^2)^2*dy-2*y*z^2,
          (x^3-y^2*z^2)^2*dz-2*y^2*z);
```

---

<sup>2</sup><http://www.math.uiuc.edu/Macaulay2>

```

II=inw(I,{0,0,0,1,1,1});
print(dim II); --- the output 4 implies that it is not holonomic.
J=WeylClosure I;
print(toString(J));
JJ=inw(J,{0,0,0,1,1,1});
print(dim JJ); --- the output 3 implies that it is holonomic.

```

We close this appendix with introducing the integration ideal. The next fact is the fundamental fact for holonomic ideals and integrations.

**Theorem 8** *If  $I$  is a holonomic ideal, then the integration ideal  $(I + \partial_d D) \cap D_{d-1}$  is a holonomic ideal in  $D_{d-1}$ . Here  $D_{d-1} = \mathbf{C}\langle x_1, \dots, x_{d-1}, \partial_1, \dots, \partial_{d-1} \rangle$ .*

This theorem follows from the fact “if  $D/I$  is a holonomic  $D$ -module, then  $D/(I + \partial_d D)$  is a holonomic  $D_{d-1}$ -module”. As to a proof of this fact, see, e.g., the Chapter 1 of the book “J. E. Björk, *Rings of Differential Operators*. North-Holland, New York, 1979”.

Oaku’s algorithm [10] to find integration ideals is explained in the Chapter 5 of [15] in a form relevant to our applications. We note that integration algorithms ([9], [10]) in  $D$  use non-term orders (see, e.g., [15, Chapter 1]). Modifications of this algorithm [9] is used in the step 1 of our Algorithm 2.

**Example 8** Put  $f(x, t) = \exp(xt - t^3)$ . The function  $f$  is annihilated by the operators  $\partial_t - (x - 3t^2)$ ,  $\partial_x - t$ , which generate a holonomic ideal  $L$ . This is a Risa/Asir code to find the integration ideal  $(L + \partial_t \mathbf{C}\langle x, t, \partial_x, \partial_t \rangle) \cap \mathbf{C}\langle x, \partial_x \rangle$ .

```

import("nk_restriction.rr");
def step1() {
  L=[dt-(x-3*t^2),
    dx-t];
  I=nk_restriction.integration_ideal(L,[t,x],[dt,dx],[1,0] | inhom=1);
  return I;
}
step1();

```

We write this introductory exposition with a few overlaps with [15]. For other fundamental facts, please refer to [15] and its references.

## References

- [1] M. Apagodu, D. Zeilberger, Multi-variable Zeilberger and Almkvist-Zeilberger algorithms and the sharpening of Wilf-Zeilberger theory, *Advances in Applied Mathematics* **37** (2006), 139-152.
- [2] O. Barndorff-Nielsen, *Information and Exponential Families*, 1978, John Wiley & Sons.
- [3] K. M. Creer, E. Irving, A. E. M. Nairn, Paleomagnetism of the great whin sill, *Geophysical Journal of the Royal Astronomical Society* **2** (1959), 306–323.

- [4] D. Cox, J. Little, D. O'Shea, *Ideals, Varieties, and Algorithms*, Third Edition, 2007, Springer.
- [5] H.Fujita, H.Konno, K.Tanabe, *Optimization methods* (in Japanese), Iwanami applied mathematics series **15**, 1998, Iwanami.
- [6] J. T. Kent, The Fisher-Bingham Distribution on the Sphere, *Journal of the Royal Statistical Society. Series B* **44** (1982), 71–80.
- [7] K. V. Mardia, P. E. Jupp, *Directional Statistics*, 2000, John Wiley & Sons.
- [8] Y. Nakamura, Gradient systems associated with probability distributions, *Japan Journal of Industrial and Applied Mathematics* **11** (1994), 21–30.
- [9] H. Nakayama, K. Nishiyama, An algorithm of computing inhomogeneous differential equations for definite integrals, [arXiv:1005.3417](https://arxiv.org/abs/1005.3417)
- [10] T. Oaku, Algorithms for  $b$ -functions, restrictions, and algebraic local cohomology groups of  $D$ -modules, *Advances in Applied Mathematics* **19** (1997), 61–105.
- [11] T. Oaku, Y. Shiraki, N. Takayama, Algebraic Algorithms for  $D$ -modules and Numerical Analysis, Z.M.Li, W.Sit (editors), *Computer Mathematics*, World scientific, 2003 (Proceedings of the sixth Asian symposium), 23–39,
- [12] T. Oaku, N. Takayama, W. Walther, A localization algorithm for  $D$ -modules. *Journal of Symbolic Computation* **29** (2000) 721–728.
- [13] K. Ohara, the yang package, which is included in the library of Risa/Asir.
- [14] Risa/Asir, a computer algebra system obtainable from <http://www.math.kobe-u.ac.jp/Asir>
- [15] M. Saito, B. Sturmfels, N. Takayama, *Gröbner Deformations of Hypergeometric Differential Equations*, 2000, Springer.
- [16] A. Snyman, *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*, 2005, Springer.
- [17] N. Takayama, Gröbner basis and the problem of contiguous relation, *Japan Journal of Applied Mathematics* **6** (1989), 147–160.
- [18] N. Takayama, An approach to the zero recognition problem by Buchberger algorithm. *Journal of Symbolic Computation* **14** (1992) 265–282.
- [19] A. T. A. Wood, Some notes on the Fisher-Bingham family on the sphere, *Communications in Statistics, Theory and Methods* **17** (1988), 3881–3897.
- [20] <http://www.math.kobe-u.ac.jp/OpenXM/Math/Fisher-Bingham.toc.html>, [fb-demo-0.txt](#), [fb-demo-1.txt](#).

Tomonari Sei, Akimichi Takemura <sup>#</sup>  
Department of Mathematical Informatics  
Graduate School of Information Science and Technolgy, University of Tokyo  
Bunkyo, Tokyo, 113-0033, Japan

Nobuki Takayama<sup>3#</sup> ([takayama@math.kobe-u.ac.jp](mailto:takayama@math.kobe-u.ac.jp)),  
Hiromasa Nakayama<sup>#</sup>, Kenta Nishiyama<sup>#</sup>, Masayuki Noro<sup>#4</sup>  
Department of Mathematics, Kobe University  
Rokko, Kobe, 657-8501, Japan

Katsuyoshi Ohara<sup>5</sup>  
Faculty of Mathematics and Physics, Kanazawa University  
Kakuma-machi, Kanazawa, 920-1192, Japan

---

<sup>3</sup>Supported by Kakenhi 19204008

<sup>4</sup> Authors with <sup>#</sup> belong to the JST crest Hibi project.

<sup>5</sup>Supported by Kakenhi 22540179